**Weather Data Classification**

## TEAM: BRILLIANCE

## 

### Abstract:

This section of the website deals with the earth's atmosphere, and particularly with the science of meteorology. Meteorology has a theoretical aspect, a branch of classical physics dealing with the statics, dynamics and thermodynamics of the moist air that makes up the atmosphere. It also has a practical aspect, weather analysis and forecasting, which is of great practical utility to agriculture, navigation and aeronautics.

**Requirements :**

The sample cleaned dataset given has text comment about weather data that needs to be classified as ‘Sunny’ or ‘Rainy’ by the application. Create a machine learning model or algorithm using Support Vector Machines technique that can be used to classify the weather for new data set.

The tasks for implementing the project are

1. Prepare the data
2. Create and train the model
3. Predict with new data
4. Calculate the accuracy of classification

**Software/Hardware requirements to implement the project:**

Hardware

* 1. Operating System – Windows/Linux (Ubuntu)
  2. Minimum 4 GB RAM
  3. 500 GB Hard disk

Software: 1. R  (<https://cran.r-project.org/>)

2. R Studio (<https://www.rstudio.com/products/rstudio/download/>)

Design:

The two classes defined for this project are C = {Sunny, Rainy}. Using a learning method or learning algorithm like Support Vector Machine, create a classifier or classification function that maps documents/text to classes. There are other techniques also that can be used to create a classifier.

The steps for implementing the project are

1. Prepare the data: we created a csv file using excel which contains a set of data inputs.
2. Create and train the model: we train the model for the set of data inputs given.
3. Predict with new data: we tried to predict the new set of data with trail and error method.
4. Calculate the accuracy of classification: the accuracy was checked for the given set of data.
5. Obtaining the result:

predictionData <- list("sunny sunny sunny rainy rainy", "rainy sunny rainy rainy", "mist sunny","fog rainy", "rainy rainy")

results:

SVM\_LABEL SVM\_PROB

1 1 0.9376577

2 -1 0.9918153

3 1 0.9377779

4 -1 0.9084324

5 -1 0.9918153

*Code:*

dataDirectory <- "H:/R/"

data <- read.csv(paste(dataDirectory, 'sunnyData.csv', sep=""), header = TRUE)

#data

dtMatrix <- create\_matrix(data["Text"])

container <- create\_container(dtMatrix, data$IsSunny, trainSize=1:45, virgin=FALSE)

# train a SVM Model

model <- train\_model(container, "SVM", kernel="linear", cost=1)

predictionData <- list("sunny sunny sunny rainy rainy", "rainy sunny rainy rainy", "mist sunny","fog rainy", "rainy rainy")

trace("create\_matrix", edit=T)

predMatrix <- create\_matrix(predictionData, originalMatrix=dtMatrix)

predSize = length(predictionData);

predictionContainer <- create\_container(predMatrix, labels=rep(0,predSize), testSize=1:predSize, virgin=FALSE)

results <- classify\_model(predictionContainer, model)

results

Conclusion & Challenges :

Weather affects every aspect of our lives, either directly or indirectly. The ability to forecast the weather accurately is an increasingly important part of our economy and our society. Weather forecasting is enormously sophisticated and expensive, involving the use of satellite technology, supercomputers and phenomenally large streams of data which are continuously being updated.

References :

Hed-X Company

• R Tutorial – TutorialsPoint – www.tutorialspoint.com/r/

• Q & A Forum – Seek help at StackOverflow – a forum for questions and answers which has rich repository of questions and solutions on R programming

Guided by :

• Nirupama Mane

• Shreshta Anantha